

Remarks

The present amendment responds to the Official Action dated June 8, 2004. The Official Action rejected claim 1 under 35 U.S.C. § 102(e) based on Gusack U.S. Patent No. 6,356,897 ("Gusack"). The Official Action rejected claims 20 and 21 under 35 U.S.C. § 102(e) based on Egan et al. U.S. Patent No. 6,356,888 ("Egan"). These grounds of rejection are addressed below following a brief discussion of the present invention to provide context. The Official Action further objected to claims 2-19 as being dependent upon a rejected base claim, and indicated that such claims would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims. Claims 1, 2, 6 and 18 have been amended to be more clear and distinct. Claims 20 and 21 have been canceled without prejudice. Claims 1-19 are presently pending.

The Present Invention

Electronic documents such as Word® documents, for example, commonly contain tabular data. Such data are electronically stored, and can be generally queried by full text searching of a Word® document. However, structured queries based on the arrangement of the tabular data and based on addressing subject headers of the tabular columns and rows, generally cannot be performed. Such tabular data, although electronic, are not stored in the form of a structured database, or otherwise in an array of fields that can be specifically queried. Reformulation of such data tables into a structured form that can be so queried is desirable, so that the location of data with respect to column headers and row headers can be determined. Such reformulation can be particularly important in the case of large volumes of tabular data. Facilitating such reformulation on electronically tabular data regardless of the table array size and shape, and

despite the absence of orderly or standardized columns, the absence of column and row dividers, the absence of boxes around headers, and the absence of other aids to defining the table structure, is also desirable. Specification, pages 1-3.

The present invention provides methods for carrying out such reformulations. In one embodiment according to the present invention, a method is provided for recognizing the structure of a delineated table region in an electronic document. Exemplary Fig. 1 shows table 10 including data spatially arranged in a delineated table region of an electronic document. The table 10 comprises data cells (Dcells) 12 and access cells (Acells) 14. These cells are organized into rows 16 and columns 18, either or both of which may or may not include headers. Table 10 includes two primary column headers in box region 22 shown as "Todays" and "Yesterdays", each overlying a pair of secondary column headers labeled "Open" and "Change", collectively designating a total of four columns. Table 10 includes seven rows of Dcells collected in a body 20 underneath the column headers in box region 22. The seven rows of Dcells are identified by row headers 24 underneath a column header in box/stub 26. Although the spatial arrangement of the data in four columns and seven rows can be plainly seen in Fig. 1, the table 10, as electronically input, lacks hierarchical arrangement sufficient to enable query of the table 10 based on such spatial arrangement. For example, it is not possible with the data, as electronically input, to select row 16 having the Acell header "Red, Inc.", and to query the table 10 as to "today's open" or as to the highest or lowest recorded price. Specification, pages 3-5.

Next, referring to Fig. 2, a binary tree 28 is created using a hierarchical clustering of a plurality of words included in the table 10. A multitude of leaves 30 constituting unique clusters and also referred to as words, are generated from the raw data of the delineated table region. The individual leaves 30 of each of these clusters are merged to form the next higher node of the

hierarchical tree. For example, at the lowest level, leaves 32 and 34 are merged to form a new cluster at higher level node 36. Similarly, leaves 38 and 40 are merged to form a new cluster at higher level node 42. At the next higher level nodes just created, cluster 36 is merged with original leaf 44 to form a second new cluster at node 46, and so on. The cluster tree 28 generated represents the hierarchical structure of the table body 20 in terms of the vertical grouping of words 30. The inter-cluster distances become grouped according to similar distances among the positional vectors of words and groups of words. The distances can be, for example, spatial, syntactic, or semantic. Specification, pages 6-9.

Next, referring again to exemplary Fig. 2, a plurality of table columns are segregated using a breadth-first traversal algorithm. A reverse process of column cutting is applied, starting at the root 50 of the cluster tree 28. It is determined that node 54 can be split into two columns, and that nodes 56, 58 and 62 cannot be so split. Hence the column cut is carried out at 48. Specification, pages 9-11.

Column headers, if any, are then identified using a first heuristic algorithm. Referring to exemplary Fig. 3, box region 22 is defined by sorting the columns according to horizontal starting position, defining the upper boundary of the table body 20 constituting the lower boundary of the box region 22, and extracting the column header information included in box region 22 using a phrase segmentation process. A tree 70 of the column headers is then created in a bottom-up manner. Specification, pages 11-15.

Row headers, if any, are then identified using a second heuristic algorithm. Next, at least one table row is segregated using a row determination algorithm. The rows 16 are accordingly defined, including identification of first "core" lines of each row and any secondary lines that follow. Specification, pages 15-16.

When the table reformulation is completed, the electronic structure is stored and can be queried using a variety of means. Specification, pages 16-21.

In another exemplary embodiment as defined in claim 1, a method is provided for recognizing the structure of a delineated table region in an electronic document, comprising the steps of: (a) inputting tabled data spatially arranged in a delineated table region of an electronic document, said tabled data, as input, lacking hierarchical arrangement sufficient to enable logical query of said tabled data based on said spatial arrangement; (b) creating a binary tree using a hierarchical clustering of a plurality of words included in said table region; (c) segregating a plurality of table columns using a breadth-first traversal algorithm; (d) identifying column headers, if any, using a first heuristic algorithm; (e) identifying row headers, if any, using a second heuristic algorithm; and (f) segregating at least one table row using a row determination algorithm.

Indication of Allowability

Applicants acknowledge with appreciation the indication that claims 2-19 are allowable. Non-substantive typographical errors in claims 2, 6 and 18 have been corrected.

The Rejection of Claim 1

Claim 1 was rejected under 35 U.S.C. § 102(e) based on Gusack. Applicants respectfully traverse this rejection and request that it now be withdrawn, in view of the above amendments and the discussion herein.

Claim 1 has been amended to recite, as its first step, that tabled data are input which are spatially arranged in a delineated table region of an electronic document, the tabled data, as so

input, lacking hierarchical arrangement sufficient to enable logical query of the tabled data based on the spatial arrangement. The preamble of claim 1, which has not been amended, recites a method for recognizing the structure of a delineated table region in an electronic document. The purpose of this amendment is to expressly include the requirement of "recognition" of such a hierarchical arrangement in the body of the claim.

The tabled data as recited in the first step of claim 1 are electronic, but they are not stored in the form of a structured database, or otherwise in an array of fields that can be specifically queried. According to the present invention as defined in claim 1, such input tabled electronic data are subjected to a series of steps that reformulates such data into a structured database that can be so queried. This method, for example, addresses the need for such reformulations as applied to such electronic tabled data which are generated by various software applications such that they may or may not contain border lines, a fixed number of blank spaces between columns, multi-line rows, multi-line column headers, or a clearly vertical column definition due to skewing. This method can be universally applied for detecting table structures, reformulating tabled data into a structured form, and enabling structured queries in tabled data as found in such software applications.

Gusack fails to disclose and fails to suggest a method step of inputting tabled data which are spatially arranged in a delineated table region of an electronic document, the tabled data, as so input, lacking hierarchical arrangement sufficient to enable logical query of the tabled data based on the spatial arrangement. To the contrary, Gusack discloses database structures into which data are then input and subjected to various linking and query procedures. Gusack fails to disclose and fails to suggest any method for reformulating electronic tabled data that are not stored in a prior-generated structured database that can be queried.

Gusack is entitled "Associative Database Model for Electronic-Based Informational Assemblies". Gusack summarizes an object of providing relational database structures as follows:

Accordingly, it is an object of this invention to provide a unique indexing system for an assembly of informational items stored on electronic-based media, constructed from at least one registration table and set of program instructions that assigns at least one unique alphanumeric indicum to each data table included in the database structure and, therefore provides a means for calculating at least one unique domain of alphanumeric indicia for each table registered in the data set, in turn, providing a means for calculating at least one unique indicia for each record and each field defined within each record in the entire data set, and therefore providing a higher degree of order, integrity, continuity, and user convenience in accessing and creating knowledge from the assembly of informational items stored in said data tables than data structures of prior art. Gusack, col. 3, lines 43-57; see also, col. 3, lines 30-36.

Gusack starts with a database structure, to which data are added and then manipulated. The Official Action makes reference to Fig. 9 of Gusack and its disclosure at col. 15, lines 27-53. Fig. 9 and the cited passage of Gusack disclose relational database structures. These citations are part of Gusack's disclosure of an embodiment of an indexing system, beginning at col. 11, line 51. The indexing system assigns to each record in a plurality of data tables, an alphanumeric indicum that is unique for all records stored in the data set. Col. 11, lines 51-54. Referring to Fig. 6 of Gusack, the first data table (601) is a table registry containing a TIN field (603) for

storing a unique alphanumeric indicum. Gusack, col. 12, lines 6-10. Hence this data table, with which Gusack's indexing system begins, is itself a relational database, capable of assigning and storing Gusack's unique alphanumeric indicum. Referring to Fig. 6, each of the other four data tables 619, 625, 631 and 637 includes an L# field 621, 627, 633 and 641, respectively, for storing unique record identifiers. Gusack, col. 12, lines 55-57. Gusack's five data tables shown in Fig. 6 clearly do not lack hierarchical arrangement sufficient to enable logical query of the tabled data based on its spatial arrangement. The unique record identification system ("URIS") is then generated by this indexing. Col. 13, lines 38-40. Once the data set is stored to provide the unique record identification indicia, a large number of options become available to structure relational links between records stored in tables included in the data set. Col. 13, lines 57-62. The actual links thus generated are then removed from each record and located in a separate set of specialized linking tables, constituted by the central linking table system ("CLTS"). Gusack, col. 14, lines 3-24. The Gusack passage on col. 15 cited by the Official Action states, "[t]he CLTS and URIS work together to allow a user to view linked records in said user interface." This discussion, and Fig. 9 to which it refers, address a relational database.

Gusack accordingly fails to disclose and fails to suggest, at col. 15, in Fig. 9 or elsewhere, inputting tabled data spatially arranged in a delineated table region of an electronic document, said tabled data, as input, lacking hierarchical arrangement sufficient to enable logical query of said tabled data based on said spatial arrangement. Applicants therefore respectfully traverse the assertions on page 2 and 3 of the Official Action that Gusack discloses a binary tree, a row determination algorithm, a breadth first traversal algorithm, and first and second heuristic algorithms, all as defined in claim 1.

Gusack further fails to disclose and fails to suggest a step of creating a binary tree from such tabled data using a hierarchical clustering of a plurality of words included in said table region. Gusack additionally fails to disclose and fails to suggest a step of segregating a plurality of table columns in such tabled data using a breadth-first traversal algorithm. Gusack further fails to disclose and fails to suggest a step of identifying column headers, if any, using a first heuristic algorithm, as defined in claim 1. Gusack additionally fails to disclose and fails to suggest a step of identifying row headers, if any, using a second heuristic algorithm, as defined in claim 1. Gusack additionally fails to disclose and fails to suggest a step of segregating at least one table row using a row determination algorithm, as defined in claim 1.

The Rejection of Claims 20 and 21

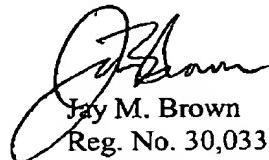
Claims 20 and 21 were rejected under 35 U.S.C. § 102(e) based on Egan. Claims 20 and 21 have been canceled. Applicants respectfully traverse this rejection and accordingly request that it now be withdrawn as moot.

Egan generally relates to a database management system performed by computers, and more specifically relates to the optimization of structured query language (SQL) queries using an encoded vector index (EVI) to process a DISTINCT function. Egan employs computer-implemented routines to query information from a database. Egan, col. 1, lines 13-17; and col. 2, lines 60 and 61. Egan accordingly fails to disclose and fails to suggest a step of inputting tabled data spatially arranged in a delineated table region of an electronic document, said tabled data, as input, lacking hierarchical arrangement sufficient to enable logical query of said tabled data based on said spatial arrangement.

Conclusion

All of the presently pending claims, as amended, appearing to define over the applied references, withdrawal of the present rejections and prompt allowance are requested.

Respectfully submitted,



Jay M. Brown

Reg. No. 30,033

Priest & Goldstein, PLLC
5015 Southpark Drive, Suite 230
Durham, NC 27713-7736
(919) 806-1600